Exploring Phonetic Category Structure with Markov Chain Monte Carlo

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with research distinction in Psychology in the undergraduate
colleges of The Ohio State University

by

James Pooley

Project Advisors: Professors Mark Pitt and Jay Myung, Department of Psychology

Abstract

Research in cognitive psychology on how humans mentally represent phonetic categories has demonstrated that these categories have a graded internal structure, meaning that some exemplars of the category are considered subjectively better, or more representative, members of the category than are others. Since many phenomena in speech perception are sensitive to this graded category structure, the determination of this structure is a goal of research in the area. The goal of this thesis is to apply a new experimental methodology to explore the mental representations of a phonetic category. The procedure exploits a connection between human choice behavior and Markov chain Monte Carlo (MCMC) algorithms to sample from an individual's mental representation of a phonetic category, taken here to be a probability distribution over speech sounds. Stimuli for the project consisted of 81 computer synthesized examples of the /i/ phonetic category. A computer program repeatedly presented pairs of these stimuli to subjects in accordance with the requirements of MCMC algorithms, and subjects' choices mimicked an MCMC acceptance function. The exemplars of /i/ chosen by subjects during the course of the experiment were used to estimate their mental representations of the /i/ phonetic category. The phonetic category structures estimated by the psychological MCMC procedure show that subjects tended to choose stimuli at the edge of the acoustic stimulus space as the best examples of the /i/ phonetic category. This result has not always been obtained with similar stimulus sets and a goodness rating technique, and some implications for future use of the procedure are discussed.

Exploring Phonetic Category Structure with Markov Chain Monte Carlo

The question of how people mentally represent various categories has played a prominent role in many areas of cognitive psychology over the years. Since these mental representations are the objects upon which numerous cognitive processes are purported to operate, their study has occupied a central place in the discipline. Early research on category representation focusing on the mental representation of natural object categories (e.g., animals, foods, etc.) found, among other things, that the mental representations of these categories have a graded internal structure; that is, despite all being identified as exemplars of the same category, certain exemplars of a given category are considered more representative of it than are other exemplars (e.g., Mervis & Rosch, 1981). As an example, consider the category BIRD. When asked to rate the representativeness of this category of various exemplar birds, subjects will predictably rate a robin as more representative of the category than a penguin. In addition to this *perceptual* gradedness in terms of perceived representativeness of the category, many of these categories have also been found to be *functionally* graded. For example, certain members of the category (e.g., the robin) are more easily remembered or named than are others (e.g., the penguin) in tasks that require subjects to remember or name a list of exemplars in response to a category prompt (e.g., BIRD).

As in these higher-level cognitive processes, mental representations of categories have also been implicated to play a role in the perception of speech. Simplifying things, the basic idea is that the continuous acoustic signal produced by a talker arriving at the ear of a listener is somehow transformed in the mind of the listener into increasingly more abstract representations of the original input during the process of speech perception, ultimately leading to the perception of the intended message. One of these transformations that is of particular interest to researchers in speech perception is thought to segment the continuous signal of acoustic input by mapping portions of it onto more abstract *phonetic category* representations, which correspond to the speech sounds used in one's native language. Assuming that some such transformation occurs during the process of speech perception, a question immediately arises: What is the nature of the mental representation of these phonetic categories, and do they too exhibit a graded category structure?

Early research on this question focused on the phenomenon of categorical perception, stressed the important role of the boundary between neighboring phonetic categories in the perception of speech, and appeared to demonstrate that categories had no graded internal structure—all exemplars were thought to be considered equally representative by subjects (e.g., Liberman, Harris, Hoffman, & Griffith, 1957). Despite these early reports, subsequent work has convincingly demonstrated that the mental representations of phonetic categories too have a graded internal structure (e.g., Kuhl, 1991; Miller, 1994; Miller, 2001; Miller & Volaitis, 1989), with exemplars of a given phonetic category differing in both their functionality and perceived representativeness of the category. Furthermore, as is the case with other psychological categories, there is a systematic relationship between these two aspects of graded category structure, and the recognition of this has fueled much research in speech perception over the past few years.

Of all the perceptual phenomena associated with the various exemplars of phonetic categories that differ in their perceived representativeness of the category, one of the most thoroughly studied has been the perceptual magnet effect for speech reported by Kuhl (1991; see also Iverson & Kuhl, 1995, 1996, 2000). The perceptual magnet effect is characterized by reduced discrimination sensitivity to stimuli near the best exemplars of the phonetic category and increased discrimination sensitivity near poor exemplars of the category; that is, there is a "warping" of perceptual space near the good exemplars. Many of the studies that have examined the perceptual magnet effect have done so in the context of both first and second language acquisition (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Kuhl, 2000, 2004). Among other things, these studies have shown that acquiring phonetic categories early in life affects the way in which languages are learned and perceived later in life (e.g., consider the difficulty many native speakers of Japanese have in acquiring the American English /r/-/l/ distinction). Since the perceptual magnet effect and other similar phenomena depend on the graded structure of phonetic categories, a crucial step in studies examining these phenomena is the determination of this structure. Currently, this has been accomplished using a variety of complimentary methods.

One way to uncover the structure of a phonetic category is to see how its various exemplars function in different experimental tasks, and from this infer the structure of the category. For example, experiments using the selective adaptation procedure have demonstrated that not all exemplars of a phonetic category are equally good adaptors in the procedure, with the perceived best exemplars of a phonetic category being the best adaptors (Miller, Connine, Schermer, & Kluender, 1983; Samuel, 1982). Another way to uncover the structure of a phonetic category is to represent the perceived distances between the category exemplars in a psychologically meaningful metric space using multidimensional scaling algorithms (Iverson & Kuhl, 1995, 1996; Kewley-Port & Atal, 1989). These algorithms construct such spaces based on pairwise ratings of (dis)similarity between the exemplars, usually obtained by either having subjects directly judge the (dis)similarity between the exemplars or indirectly by using confusion data. By far, however, the most common and direct method used to uncover the structure of the mental representation of a phonetic category is simply to ask individuals to rate the perceived “goodness” of the exemplar as a member of the category to which it belongs on a numeric scale.

The goodness rating procedure is widely used to investigate phonetic category structure, but its use is problematic for a number of reasons. Chief among these reasons is that, due to the ordinal nature of goodness ratings (Stevens, 1946), all that can be inferred from a subject's responses is a preference ordering. For example, if a subject gives one exemplar a goodness rating of 5 and another 7, we cannot conclude that the subject considers the difference in perceived goodness between these two exemplars to be twice that of an exemplar receiving a rating of 5 and another that receives a rating of 6. Thus, goodness ratings can at best give only a rough estimate of phonetic category structure. Additionally, it is unclear what biases this procedure is inducing in the subjects. When using goodness ratings, subjects are instructed to take the whole set of category exemplars they are rating into account and to keep track of the ratings they gave to previous exemplars when giving a rating to any single exemplar. It is quite possible that subjects do not follow this instruction, which could influence the ratings in unseen ways. Thus, it seems that there room for considerable methodological improvement in the area of uncovering the representation of phonetic categories.

Recently, a novel experimental procedure for uncovering the structure of the mental representations of categories has been developed by Sanborn (2007; see also Sanborn & Griffiths, in press). The method avoids many of the aforementioned shortcomings that make the other techniques currently used for estimating phonetic category structure problematic by exploiting the strengths of a powerful and popular statistical sampling method within a psychological context. Markov chain Monte Carlo (MCMC), the class of sampling methods upon which the new experimental procedure is based, is a standard method used in many disciplines to efficiently draw samples from complex probability distributions. Before discussing how the psychological application of MCMC works and the assumptions that lead to this procedure, a general overview of the very basics of a standard MCMC algorithm will be given to help make its use in a psychological context clear.

As mentioned previously, MCMC algorithms are used to draw samples from complex probability distributions. The algorithms accomplish this by iteratively generating two samples following a prescribed algorithmic rule such that one of the samples is the current state and the other the proposed state of a Markov chain for which the stationary distribution is the target distribution one wants to draw samples from (for an overview of the MCMC procedure, see Gilks, Richardson, & Spiegelhalter, 1996). A typical MCMC algorithm works as follows: After arbitrarily initializing the current state of a Markov chain, a proposal stimulus is generated from a proposal distribution which is dependent[1] on this current state of the Markov chain. An acceptance function then determines the probability that the proposed state becomes the new current state of the chain. The current state of the Markov chain then changes with the probability given by the acceptance function, and the procedure is iterated until the samples produced by the procedure converge to the distribution of interest (for more detail on convergence, see Results).

Assuming that categories may be considered probability distributions over exemplars (an assumption often used in many models in cognitive psychology; e.g., Ashby & Alfonso-Reese, 1995) and deriving a correspondence between the Luce (1963) choice rule and an acceptance function sometimes used in MCMC algorithms, Sanborn (2007) and Sanborn and Griffiths (in press) developed an experimental procedure in which a human subject is forced to act like a component of an MCMC algorithm. Specifically, the subject is made to mimic an MCMC acceptance function,[2] so that their choices influence which stimuli they are presented for comparison throughout the course of the experiment and that the accepted stimuli form the states of a Markov chain for which the stationary distribution is the psychological category of interest. The basic idea behind the psychological application of the MCMC procedure, then, is to estimate the structure of a mental representation of a category *implicitly* based on an extended series of simple two-alternative, forced-choice decisions, without having to ask individuals to give vague numerical ratings of how good the exemplars are of a given category or of how similar two exemplars are.

Given that phonetic category structure has been shown to play a role in many phenomena in speech perception, it should be evident that a representative delineation of phonetic category structure is highly desirable. The purpose of the experiment reported in this thesis is to see if the psychological MCMC procedure just described is able to provide a better characterization of phonetic category structure than is currently available from the use of other methods. To do this, I have adopted and slightly adapted the psychological MCMC procedure developed and applied by Sanborn (2007) and Sanborn and Griffiths (in press) in the context of natural object categories to estimate the structure of a phonetic category. Specifically, the experiments in this thesis will focus on the /i/ phonetic category. This phonetic category has been used in much of the research on the perceptual magnet effect, and an estimate of its structure and function has already been established by numerous experiments (e.g., Iverson & Kuhl, 1995, 1996; Kuhl, 1991). This will allow a comparison with what is obtained with the psychological MCMC procedure used here.

What I attempt to show here is that using the psychological MCMC procedure as opposed to the standard goodness rating procedure to uncover phonetic category structure has many potential benefits. One such benefit is that the psychological MCMC procedure is a powerful procedure that is designed to estimate complex distributions. Thus, the psychological MCMC procedure could potentially uncover subtle (and perhaps important) differences in phonetic category structure that cannot be observed using only ordinal goodness ratings to estimate category structure. Thus, the better characterization of the structure of phonetic categories uncovered by the psychological MCMC procedure could shed new light on various perceptual phenomena associated with the representation of phonetic categories and point the way for new research.

Method

Participants

Five undergraduate students enrolled in the introductory psychology course at The Ohio State University participated in two sessions of the experiment in exchange for course credit. All were from various regions of Ohio.

Stimuli

A set of 81 exemplars of the phonetic category /i/, similar to the set used by Kuhl (1991), was synthesized using Praat (Boersma & Weenink, 2007). The 81 exemplars were created by varying the values of F1 with each of the nine possible values for F2 and keeping F3 through F5 constant for each stimulus at the values used by Kuhl (1991). Values for F1 ranged from 191 Hz to 364 Hz, and values for F2 ranged from 1982 Hz to 2763 Hz. In accordance with earlier studies, the perceived distance between the F1 and F2 adjacent tokens was equated on a psychophysical scale; however, the equivalent rectangular bandwidth (ERB) scale (Moore, 2003, p. 74) was used in our study rather than the mel scale (Stevens, Volkmann, & Newman, 1937), which has been used in past studies with similar stimulus sets. The formants of each stimulus were separates by 0.45 ERB. Each stimulus was also given a linearly falling pitch contour from 130 to 100 Hz to make the stimulus sound more natural and stimuli were equated in RMS amplitude.

Apparatus and Procedure

Stimuli were presented to subjects in a sound-controlled booth by a script running in the stimulus control program Presentation (Neurobehavioral Systems, 2007) on a PC.

Subjects participated in four blocks of trials, separated over two successive days, with each block consisting of 300 MCMC trials for a total of 1200 such trials per subject. On each trial, subjects were presented with two of the synthesized exemplars of the /i/ vowel category, and they were instructed to choose which of the two exemplars was the better example of the vowel sound in the word “key,” which was printed on a board and pointed to rather than spoken. After the initial presentation of the trial-initial exemplar, subjects could replay either of the stimuli as many times as they needed and in whatever order they wanted in order to make their choice, or they could choose one of the two stimuli as the better /i/. Once they made a choice, the chosen stimulus was kept as one of the members of the pair of stimuli presented on the subsequent trial, and the other member was chosen from one of the 81 stimuli in the stimulus space with equal probability. Subjects were given a short practice session of 10 such MCMC trials at the start of the experiment in order to familiarize them with the task.

Results

Checking for Convergence

As mentioned in our brief introduction of MCMC algorithms, it must be checked to see if the Markov chains have converged to their stationary distribution—that is, that the states of the Markov chain being generated by the MCMC algorithm are indeed samples from the distribution of interest—before the algorithms produce good results. In a psychological context, this means that we need to check if subjects are in fact choosing exemplars in such a way that their choices reflect their mental representations of /i/. The theory of MCMC states that, as long as the procedure is set up in the correct way, then this will eventually happen. Unfortunately, there is no guaranteed way to check when this point has been reached and the procedure is producing good samples. Thus, we applied a simple heuristic to evaluate convergence: For each subject, the F1 and F2 values of the exemplar they chose on each trial was averaged over the four blocks of the experiment, and it was visually inspected to see when and/or if the graph these averages roughly stabilized. The rationale behind this heuristic is that, if the subject is indeed choosing exemplars that reflect an underlying mental representation of /i/, then they should be choosing exemplars in roughly a specified subset of the stimulus space rather than continually exploring the entire space. As an added check to give some baseline of what to expect if a subject was *not* choosing exemplars in accordance with the assumptions needed for the psychological MCMC method, this procedure was also applied to a pseudosubject, which was simply a computer program that randomly chose between the two exemplars it was presented on each trial with probability 0.5.

The results of these analyses are shown in Figure 2. One noticeable aspect of these is that, the pseudosubject produces a qualitatively different pattern than the five true subjects (Figure 2a). While the F1 and F2 averages for the true subjects show periods of relative stability (Figures 2b-f), the averages for the pseudosubject never seem to settle into stable pattern. It should also be noted that, for all subjects, the F2 averages vary more than the F1 averages. One possible reason for this difference in variability is that the differences in F2 values of the exemplars spanned a larger range than did the F1 values. Another is that the proposal distribution for choosing exemplars to present to subjects was uniform over the entire stimulus space, increasing variability. Basing convergence on the averages of the F1 values alone, it seems as if each of the true subjects more or less immediately began sampling from their mental representations of /i/. Even assessing convergence by the F2 averages alone, subjects still appear to converge from the start of the procedure (even the less stable F2 averages appeared to be regular for some subjects, and this is simply treated as intrinsic variability for the subject). For this reason, no trials were removed from the subjects' chains before evaluating the phonetic category structures estimated by the MCMC procedure.

Estimated Phonetic Category Structures

The estimated phonetic category structures produced by the MCMC method for the six subjects are shown in Figure 3. In the figures, each circle represents an exemplar of /i/ and the size of the circle corresponds to the number of times a subject chose that particular exemplar as the better example of /i/ when it was presented in a pair with another exemplar from the stimulus space. Based on these data, qualitative rather than quantitative evaluations were carried out, as the data clearly violate some of the assumptions (e.g., normality) that standard hypothesis-testing procedures used in psychology to analyze data require.

Figure 3a contains the estimated structure for the pseudosubject. As a cursory visual comparison with the other estimated structures makes clear (and as would be expected based on the convergence checks mentioned above), the pseudosubject behaved quite differently from the five true subjects (Figures 3b-f). While the other subjects tended to choose exemplars in a certain subset of the stimulus space and avoided exemplars with low values for both F1 and F2, the pseudosubject chose exemplars over the entire range of the stimulus space with more or less equal frequency. Thus, the subjects behaved quite differently from the true subjects.

As well as being different from the pseudosubject's estimated phonetic category structure, the data from the five true subjects all appeared qualitatively different from each other. The data from Subject 1 (Figure 3a) in particular were quite orderly, and this subject seemed to prefer exemplars with a single F1 value and increasing F2 values in the region on the left side of the stimulus space. All other subjects appeared to prefer the opposite edge of the stimulus space. One commonality that the data of all subjects shared is that the exemplar most frequently chosen was located at the edge of the stimulus space, but just what edge this was differed between subjects. Apart from the orderly data of Subject 1, there was one other noticeable trend. Subjects 2, 4, and 5 all had a similar clustering in the top right corner of the stimulus space. In general, even though some subjects choices were more spread out over the stimulus space than others (e.g., Figs. 3d vs. 3e), no subject chose values with small F1 and F2 values with great frequency or (for some subjects) even at all.

Discussion

Perhaps the most striking aspect of the results is that, when compared to results obtained in earlier studies using the same phonetic category (Iverson & Kuhl, 1995; Kuhl, 1991), my data are *much* less orderly. These previous studies used a goodness rating procedure and found structures (when averaged over subjects) that had strong central tendencies, with goodness ratings falling off in a predictable fashion as distance from this exemplar increased. In contrast, my data show that people choose exemplars on the edges of the stimulus space with the greatest frequency. If the representations estimated by the MCMC method are indeed correct, then it also seems as if subjects have mental representations of /i/ that are composed of regions where goodness increases and decreases in an orderly fashion as well as regions where the exemplars are all considered equally good or bad exemplars of the category. What could be the cause of these findings?

One possible explanation is some type of "hyperspace effect" is occurring (Johnson, 2000). The hyperspace effect is found when an individual prefers exemplars of a phonetic category that are more extreme than the values they are usually exposed to. To truly be considered a hyperspace effect, production data from the individuals would be required. While no production data were collected to support this speculation, these data are easy to obtain and the speculation is easy to test.

Another possible explanation for the discrepancy is that it is an artifact of the MCMC methodology. The more interesting conclusion, though, is that the MCMC method has uncovered what is potentially smoothed over by a goodness rating technique. Since subjects tended to choose exemplars at the edge of the stimulus space most frequently, it is possible that they would still choose exemplars with even more extreme F1 or F2 values. Thus, perhaps the data would start to look more orderly with an extended stimulus space. Given how the representations currently look, a complete and orderly category structure is unlikely, but checking this should also be relatively easy. Either the stimulus space could be extended prior to the experiment, or the procedure could be implemented in such a way that there is no predefined stimulus space and exemplars are synthesized on the fly. It must be kept in mind that extending the stimulus space makes the version of MCMC I applied less efficient, so changes in the proposal distribution would need to be considered. For example, with a stimulus set of 81 exemplars, having a uniform proposal distribution over the entire state space is not a real problem; however, if the stimulus space way increased to include, say, 169 or more exemplars, then the method as I have applied it will become increasingly more inefficient and many more MCMC trials will be required.

A third possible explanation is simply that the differences arise due to differences in dialect. The earlier studies by Kuhl (1991) and colleagues were carried out in Washington, and given the differences between the dialects of different regions of the United States, it is possible that this plays some role. While no systematic relationship was found between the regions the subjects grew up in (information obtained from a post-experiment questionnaire) and the estimated structures, given the differences in the exemplars of any given phonetic category that subjects have been exposed to during their lifetimes (or perhaps even in a single day), it does not seem too surprising that differences could possibly arise from different subject populations with different backgrounds as a result of these influences.

Another prominent finding is the large amount of individual differences present in the estimated phonetic category structures. While individual differences have been found in previous studies on phonetic categories (Iverson & Kuhl, 1996), the differences here are particularly striking. Even though all subjects reported from being from different regions of Ohio (two from the North, two from the South, and one from Central Ohio), according to the phonetic category structures estimated by the MCMC procedure, each subject had a qualitatively different mental representation on /i/. Unfortunately, as mentioned previously, I was not able to discern any interpretable relationship between the region in which a subject grew up and their mental representation of /i/. For example, even though the exemplar most frequently chosen by Subjects 1 and 2 was identical, one subject was from central Ohio while the other was from Northeast Ohio. Regardless of whether or not there is a relationship between region and representation, however, individual differences are clearly present.

Since the psychological MCMC method used here is new, some important limitations and suggested improvements should be mentioned. One current limitation of the psychological MCMC method (as well as for MCMC methods in general) is finding an appropriate way to check for convergence. The heuristic I applied is simple but not completely satisfactory. It would have been helpful if it found regions where the chains did not converge. One reason that no trials needed to be discarded before analyzing the data is that most of the stimulus space was composed of exemplars of /i/ to begin with, so subjects did not truly have to explore much of the space before finding the exemplars of /i/ that reflect their mental representations of this phonetic category. Similarly, a better method for analyzing the data returned by the psychological MCMC procedure is desirable. Currently, I simply looked for patterns during a visual inspection of the data, but it is well known that humans are good at seeing patterns that are not there, so a more objective method of evaluating the similarities and differences between the estimated category structures is needed. A third limitation is the possibility that selective adaptation effects are influencing the estimated structures. Since subjects may potentially hear a “good” exemplar many times in a row, their category boundaries may very well be shifting during the course of the experiment.

Other than the aforementioned extensions to the MCMC methodology applied here, the use of a psychophysical method such as selective adaptation could further help validate the representations it estimates. Since good exemplars of a category have been shown to be good adapters (Samuel, 1982), if the MCMC method produces an estimate of phonetic category structure that looks questionable, comparing how the exemplars in question function in an adaptation task could provide a test of the validity of the structure. For example, if the estimated phonetic category structure has two regions where exemplars are chosen frequently and a region in between with less frequency, then the exemplars in this middle region should be poorer adaptors in a selective adaptation experiment.

Regardless of whether or not any of these potential extensions to the current procedure are implemented, I have shown that the psychological MCMC method was able to estimate what I believe are reasonable phonetic category structures of /i/ for subjects. Furthermore, these estimated structures contain some prominent differences with the results obtained on similar stimulus sets using a goodness rating procedure, which may provide evidence for the view that mental representations of speech categories are not as orderly and as simple as is somtimes found and assumed.

References

Ashby, F. G., & Alfonso-Reese, L. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216-233.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer [Computer software]. Retrieved from http://www.praat.org/

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.) (1996). *Markov chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553-462.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 92*, 1130-1140.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics, 62*, 874-886.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47-B57.

Johnson, K. (2000). Adaptive dispersion in vowel perception. *Phonetica, 57*, 181-188.

Kewley-Port, D., & Atal, B. S. (1989). Perceptual differences between vowels located in a limited perceptual space. *Journal of the Acoustical Society of America,* Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

Kuhl, P. K. (2000). A new view on language acquisition. *Proceedings of the National Academy of Sciences, 97*, 11850-11857.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*, 831-843.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York, NY: Wiley.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271-285.

Miller, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal structure, context effects and speeded categorization. *Language and Cognitive Processes, 16*, 683-690.

Miller, J. L., Connine, C. N., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America, 73,* 2124-2133.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505-512.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). Academic Press.

Neurobehavioral Systems (2007). Presentation [Computer software]. Retreived from http://www.neurobs.com/

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics, 31*, 307-314.

Sanborn, A. N. (2007). *Uncovering mental representations with Markov chain Monte Carlo*. Unpublished doctoral dissertation, Indiana University: Bloomington, IN.

Sanborn, A. N., & Griffiths, T. L. (in press). Markov chain Monte Carlo with people. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems 20*.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Stevens, S. S., Volkmann, J., Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America, 8*, 185-190.

Footnotes

[1] This is not necessarily the case, as MCMC algorithms can be constructed in such a way that the proposal distribution is uniform over the entire state space and a proposed state is generated independently of the current state of the Markov chain. The algorithm described in the body of this thesis is a simple example of what is known as the Metropolis-Hastings algorithm, and the modified version we use is a modification known as the Independent Metropolis-Hasting algorithm.

[2] Under the Sanborn (2007) and Sanborn and Griffiths (in press) analysis of a simple choice task, such as the one used in the experiment reported here, Luce's (1963) choice rule gives the probability of choosing stimulus x_1 as the better exemplar of a given category as

$$p(h_1 \mid x_1, x_2) = \frac{p(x_1 \mid c)}{p(x_1 \mid c) + p(x_2 \mid c)},$$

where h_1 is the subjects' hypothesis that x_1 is the object drawn from the category of interest and x_2 is from some other distribution. If, with the correct experimental setup, x_1 is the current state of a Markov chain and x_2 the proposed state, then this is formally equivalent to the Barker acceptance function sometimes used in MCMC algorithms:

$$A(x^*; x) = \frac{p(x^* \mid c)}{p(x^* \mid c) + p(x \mid c)},$$

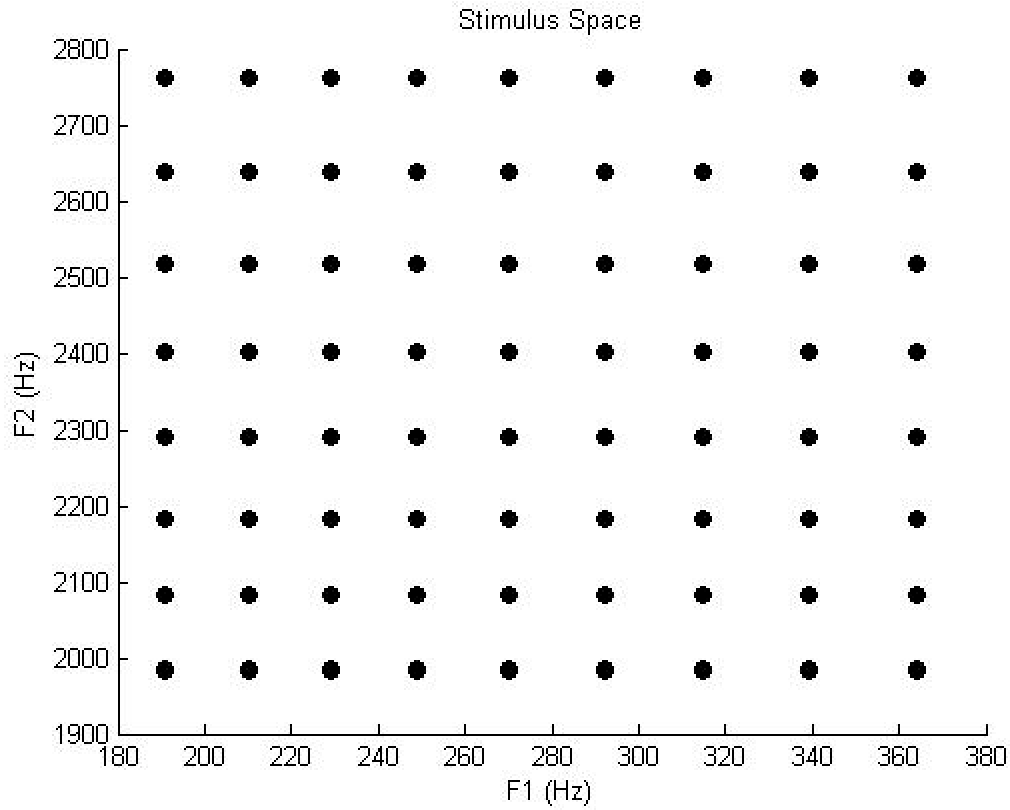
where x^* is the proposed state and x the current state of a Markov chain.

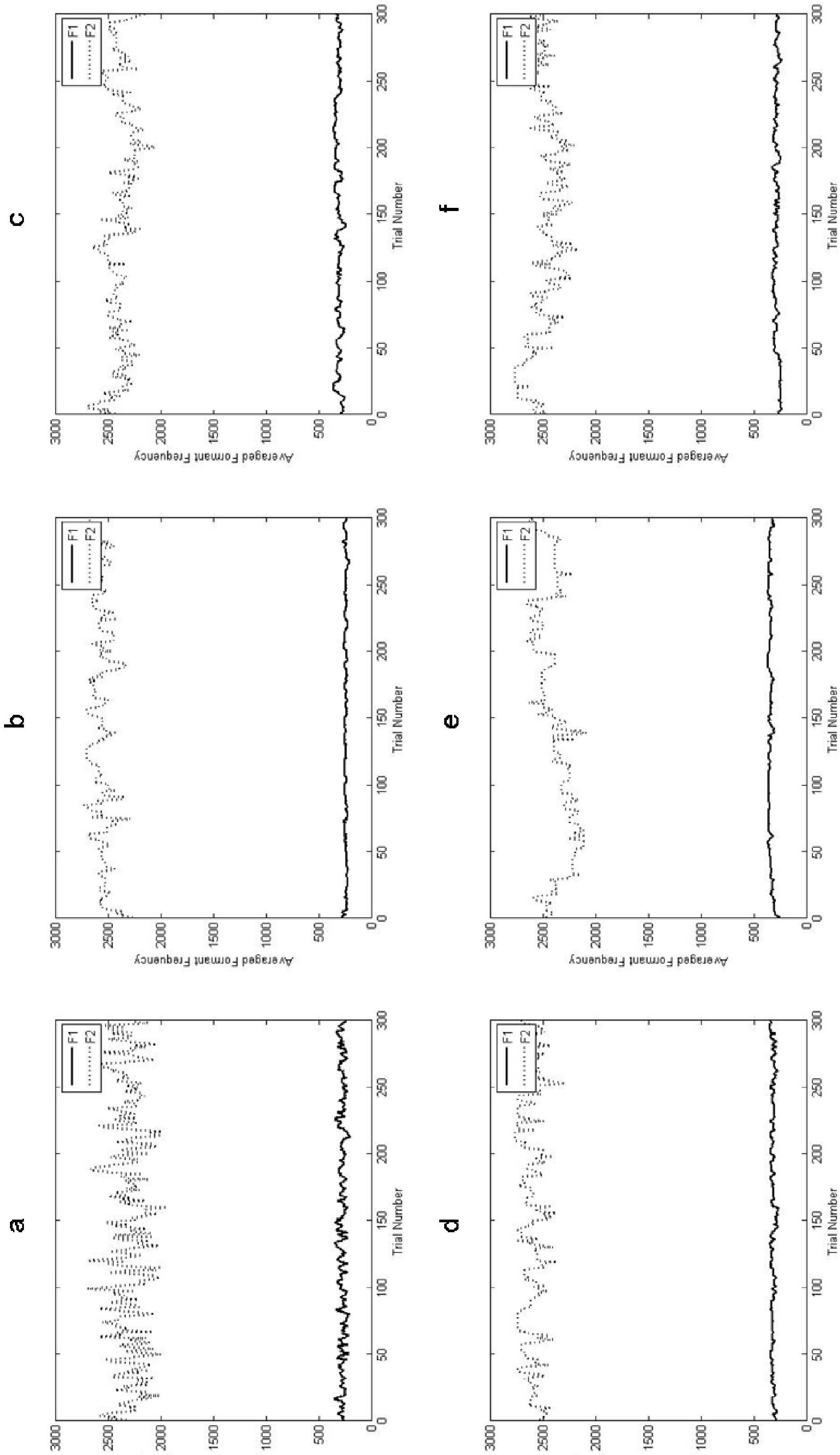
Figure Captions

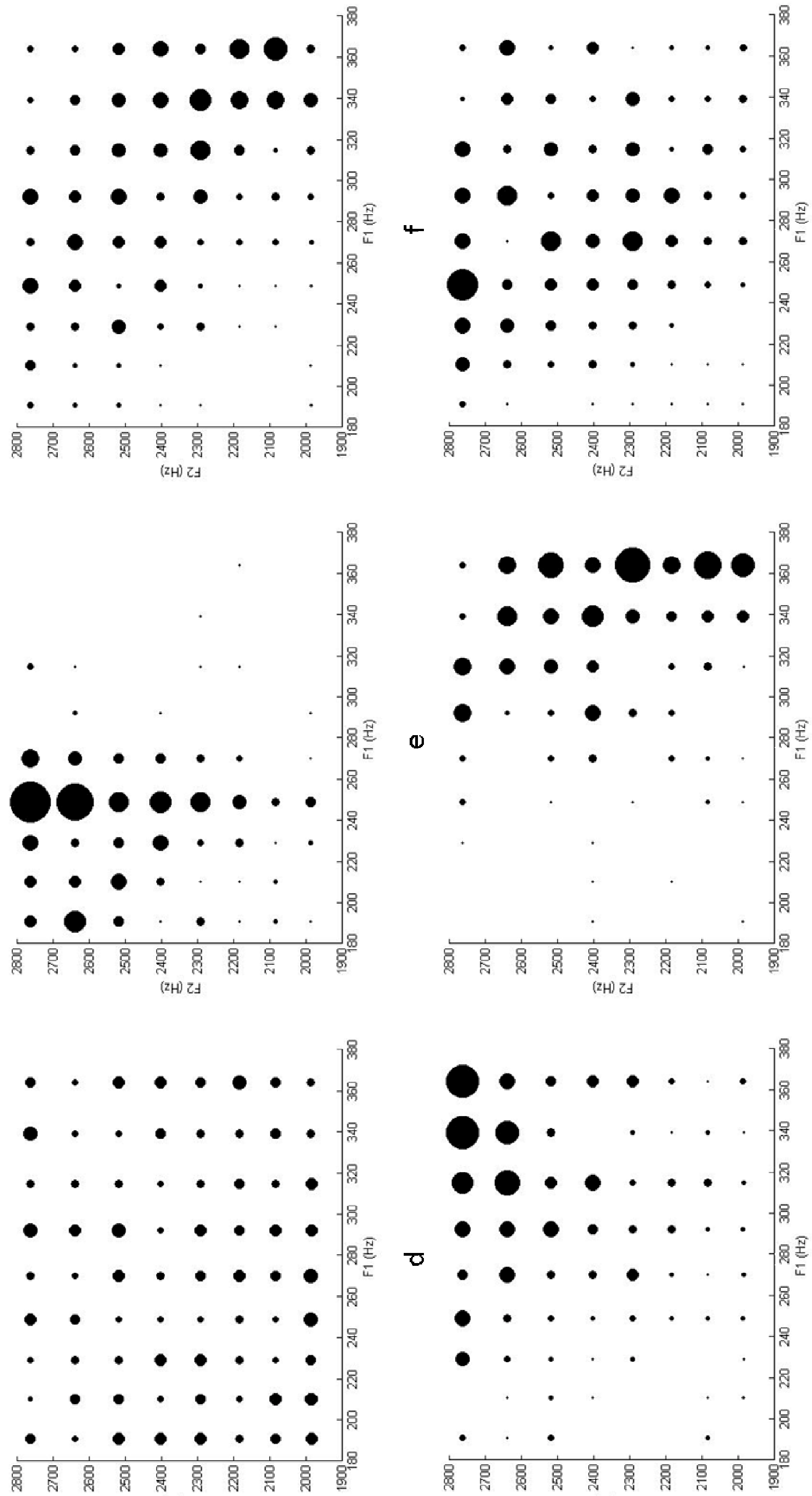
Figure 1. Stimulus space for the experiment. Each circle corresponds to an exemplar used in the experiment.

Figure 2. Averaged formant frequencies (F1 and F2) over the four blocks of the experiment used as a heuristic to check for convergence for each subject. Not the qualitatively different patterns obtained for the pseudosubject and the true subjects.

Figure 3. Phonetic category structures estimated by the psychological MCMC method for each subject. Each circle corresponds to an exemplar and the size of each circle corresponds to the number of times it was chosen. Not the qualitatively different patterns obtained for the pseudosubject and the true subjects.


Stimulus Space
F2 (Hz)
F1 (Hz)
2800
2700
2600
2500
2400
2300
2200
2100
2000
1900
180
200
220
240
260
280
300
320
340
360
380


a
b
c
d
e
f
F1
F2
Averaged Formant Frequency
Trial Number


a
b
c
d
e
f
F1 (Hz)
F2 (Hz)